Research Report

Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity

Ananthanarayan Krishnan*, Jackson T. Gandour*, Gavin M. Bidelman

Department of Speech Language Hearing Sciences, Purdue University, 1353 Heavilon Hall, 500 Oval Drive, West Lafayette, IN 47907-2038, USA

ARTICLE INFO

ABSTRACT

It has been demonstrated that neural encoding of pitch in the auditory brainstem is shaped by long-term experience with language. To date, however, all stimuli have exhibited a high degree of pitch saliency. The experimental design herein permits us to determine whether experience-dependent pitch representation in the brainstem is less susceptible to progressive degradation of the temporal regularity of iterated rippled noise (IRN). Brainstem responses were recorded from Chinese and English participants in response to IRN homologues of Mandarin Tone 2 ($T2_{IRN}$). Six different iterations steps were utilized to systematically vary the degree of temporal regularity in the fine structure of the IRN stimuli to produce a pitch salience continuum ranging from low to high. Pitch-tracking accuracy and pitch strength were computed from the brainstem responses using autocorrelation algorithms. Analysis of variance of brainstem responses to $T2_{IRN}$ revealed that pitch-tracking accuracy is higher in the native tone language group (Chinese) relative to the non-tone language group (English) except for the three lowest steps along the continuum, and moreover, that pitch strength is greater in the Chinese group even in severely degraded stimuli for two of the six 40-ms sections of $T2_{IRN}$ that exhibit rapid changes in pitch. For these same two sections, exponential time constants for the stimulus continuum revealed that pitch strength emerges 2–3 times faster in the tone language than in the non-tone language group as a function of increasing pitch salience. These findings altogether suggest that experience-dependent brainstem mechanisms for pitch are especially sensitive to those dimensions of tonal contours that provide cues of high perceptual saliency in degraded as well as normal listening conditions.

1. Introduction

There is a growing body of empirical evidence to support the notion that the neural representation of pitch may be influenced by one's experience with language (or music) at *subcortical* as well as cortical levels of processing (for reviews, see Johnson et al., 2005; Kraus and Nicol, 2005; Kraus and Banai, 2007; Patel and Iversen, 2007; Tzounopoulos and Kraus, 2009; Zatorre and Gandour, 2008). Pitch provides an excellent window for studying language-dependent effects on subcortical processing as it is one of languages' most important information-bearing components. *Tone languages*, in particular, are especially advantageous for investigating the linguistic use of pitch because variations in pitch patterns at the syllable

---

* *Corresponding authors.* Fax: +1 765 494 0771.
  E-mail addresses: rkrish@purdue.edu (A. Krishnan), gandour@purdue.edu (J.T. Gandour).

level may be lexically significant (Yip, 2003). Mandarin Chinese, for example, has four lexical tones: ma^1 "mother" [T1], ma^2 "hemp" [T2], ma^3 "horse" [T3], ma^4 "scold" [T4] (Supplementary Material: f0contours_T1-T4.doc).

As a window into pitch processing in the brainstem, we utilize the human frequency following response (FFR). The FFR reflects sustained phase-locked activity in a population of neural elements within the rostral brainstem (Alkhoun et al., 2008; Galbraith et al., 2000; Glaser et al., 1976; Krishnan, 2006, review; Marsh et al., 1974; Smith et al., 1975; Worden and Marsh, 1968). It is characterized by a periodic waveform that follows the individual cycles of the stimulus waveform. Experimental evidence overwhelmingly points to the inferior colliculus (IC) as the source of the FFR generator.

As reflected by FFRs, previous comparisons between native speakers of tone (Mandarin) and non-tone (English) languages show that native experience with lexical tones enhances pitch encoding at the level of the brainstem irrespective of speech or nonspeech context (Krishnan et al., 2005, 2009a; Swaminathan et al., 2008a). Moreover, language-dependent pitch encoding mechanisms in the brainstem appear to be especially sensitive to the curvilinear shape of pitch contours. Using trilinear and linear approximations to a natural, curvilinear pitch contour, no language-dependent effects are observed regardless of how close a *linear* pitch pattern approximates a native lexical tone (Krishnan et al., 2009a; Xu et al., 2006b). Curvilinearity itself, though necessary, is insufficient to enhance pitch extraction of the auditory signal at the level of the brainstem. A nonnative curvilinear pitch pattern similarly fails to elicit a language-dependent effect (Krishnan et al., 2009a). We therefore conclude that language-dependent neuroplasticity occurs only when *perceptually salient* dimensions of pitch in the auditory signal are part of the listener's experience and relevant to speech perception.

Most human communication occurs against a background of noise. It is important that the auditory system has a mechanism(s) in place to encode behaviorally relevant acoustic features relevant to pitch that may be degraded in the presence of noise. By using iterated rippled noise (IRN) homologues of pitch contours associated with native lexical tones, we are able to systematically vary their degree of temporal regularity and, as a consequence, their pitch salience. The question then arises whether language-dependent pitch representation in the brainstem is less vulnerable to systematic degradation in the temporal regularity of an IRN stimulus that represents a native pitch contour.

IRN stimuli preserve the perception of pitch yet do not have waveform periodicity or highly modulated stimulus envelopes that are characteristic of speech stimuli. IRN stimuli are generated using a broadband noise which is delayed and added back onto itself repeatedly. Successive iterations in this process introduce temporal regularity into the fine structure of the noise, producing a "ripple" in its long-term power spectrum. The pitch of IRN stimuli is equal to the reciprocal of the time-lag in the autocorrelation function at which the first peak occurs (Patterson et al., 1996; Yost, 1996b). The increase in the number of iteration steps systematically increases the temporal regularity in the fine structure of the noise, presumably by increasing the probability of pitch intervals at the delay, which is associated with an increase

in the pitch strength or salience of the perceived pitch. This increased pitch salience is represented by the magnitude of the autocorrelation peak. Perceptually, IRN sounds like a buzz tone with a pitch equal to the reciprocal of the delay plus some amount of background noise. The perceived tone/noise ratio (cf., SNR) increases with the number of iterations.

Using IRN stimuli, psychophysical (Patterson et al., 1996; Yost, 1996a, 1997), physiological (Bilsen et al., 1975; Fay et al., 1983; Griffiths et al., 1998; Sayles and Winter, 2007; Shofner, 1991, 1999; Winter et al., 2001, 2003), electrophysiological (Krumbholz et al., 2003; Soeta et al., 2005), and imaging experiments (Griffiths et al., 1998, 2001) have all consistently shown an increase in neural activity at the pitch period with increase in the number of iteration steps of the IRN stimulus. It is generally accepted that the pitch of the harmonic IRN is represented by neural activity phase locked to either the fine-structure or envelope periodicity.

The stimuli we choose to investigate are IRN homologues of Mandarin Tone 2 ($T2_{IRN}$), a high rising pitch contour in the tonal space. In previous cross-language comparisons (Chinese vs. English), Tone 2 ($n = 32$) yields enhanced pitch representations for native Chinese over English, especially in those portions exhibiting rapid changes in pitch regardless of context, speech, or nonspeech, i.e., IRN (Krishnan et al., 2009a,b; Swaminathan et al., 2008a). Thus, we conclude that $T2_{IRN}$ stimuli, characterized by a high degree of pitch salience, are veridical representatives of *linguistic* pitch. However, whether their superiority in brainstem pitch encoding is maintained when presented with degraded stimulus periodicities, i.e., $T2_{IRN}$ stimuli generated with lower numbers of iteration steps, is the question to be addressed herein. We expect the native Chinese to show (1) an advantage over speakers of a non-tone language even in less-than-ideal listening conditions as well as to show (2) that experience-dependent neural mechanisms for pitch representation at the brainstem level are sensitive to *specific dimensions* of pitch contours that are highly relevant in degraded as well as normal listening conditions.

2. Results

2.1. Temporal properties of whole tone

Autocorrelograms derived from the grand averaged FFR waveforms in response to IRN homologues of Tone 2 at low (top row), intermediate (middle row), and high (bottom row) iteration steps are shown in Fig. 1 for the Chinese and English groups. At low iteration steps ($n = 2$), bands of phase-locked activity are less distinct and more diffuse regardless of language group. By an intermediate number of iteration steps ($n = 8$), the autocorrelograms of the Chinese group, as compared to the English, show clearer dark bands of phase-locked activity at the fundamental period (f_0) and its multiples throughout the entire duration of the FFR response. At high iteration steps ($n = 32$), Chinese listeners' superiority is even more pronounced, as evidenced by their darker, narrower bands throughout the response. These narrower bands indicate that phase-locked activity for Chinese listeners is not only more robust but also more accurate than that of English.
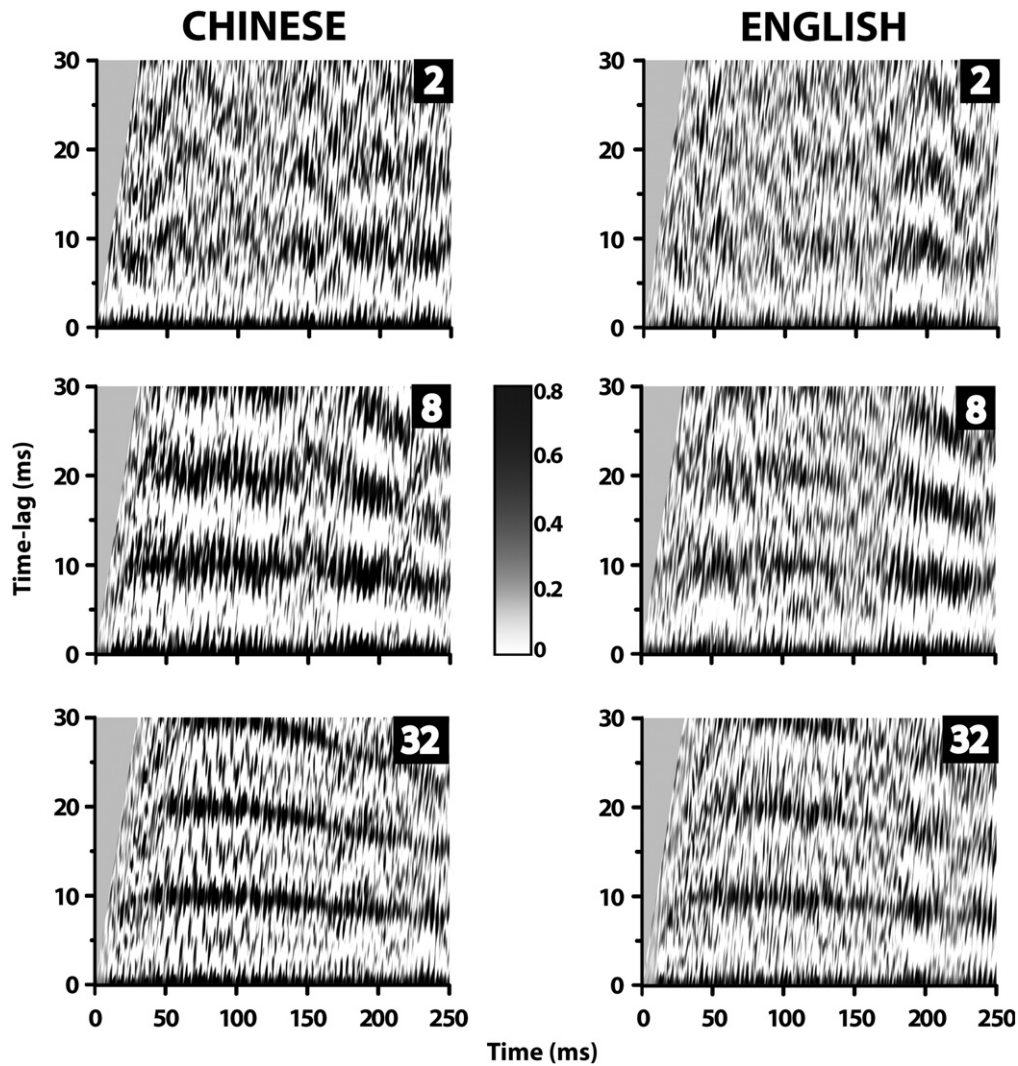
Fig. 1 – Correlograms derived from grand averaged FFR waveforms of Chinese (left) and English (right) groups in response to T2$_{IRN}$ at low ($n=2$), intermediate ($n=8$), and high ($n=32$) iteration steps. Little difference is seen between the Chinese and English group at low iteration steps ($n=2$; top row). By $n=8$ iterations, Chinese show clearer, tighter bands of temporal regularity (black) in FFR phase-locked activity at the fundamental period ($1/f_0$) and its multiples as compared to the English group (middle row). At $n=32$, this superiority is even more pronounced (bottom row). ACF magnitudes are indicated by the color gradient; darker shades indicate higher temporal regularity (i.e., more pronounced phase-locked activity). See Fig. 4 for the stimulus pitch contour. FFR, frequency following response; ACF, autocorrelation function.

2.2. Pitch-tracking accuracy of whole tone

FFR pitch-tracking accuracy, as measured by the cross-correlation between the stimulus and response pitch contours, is shown for IRN homologues of Tone 2 across iteration steps per language group (Fig. 2). Except for low iteration steps, we observe that pitch tracking is more accurate in the Chinese than the English group. This means that f_0 contours derived from the FFR waveforms by the Chinese group more closely approximate those of the original IRN stimuli.

An omnibus two-way ANOVA on cross-correlation coefficients between the stimulus (T2$_{IRN}$) and response (FFR) contours yielded significant main effects of group ($F_{1,18}=19.86$, $p=0.0003$) and iteration ($F_{5,90}=15.86$, $p<0.0001$), as well as a group×iteration interaction effect ($F_{5,90}=2.43$, $p=0.0409$). Post

hoc Tukey–Kramer pairwise comparisons of group ($\alpha=0.05$) revealed that pitch-tracking accuracy of T2$_{IRN}$ was significantly higher in the native Chinese group as compared to the English beginning as early as 12 iteration steps ($n\geq12$).

2.3. Pitch strength of tonal sections

FFR pitch strength, as measured by the magnitude of the normalized autocorrelation peak, is shown for each of six sections within IRN homologues of Tone 2 as a function of iteration steps per language group (Fig. 3).

Results from an omnibus three-way (group×section× iteration) ANOVA on pitch strength yielded significant ($p<0.0001$) main effects (group: $F_{1,18}=45.67$; section: $F_{5,90}=20.65$; iteration: $F_{5,90}=114.27$) and significant two-way interactions
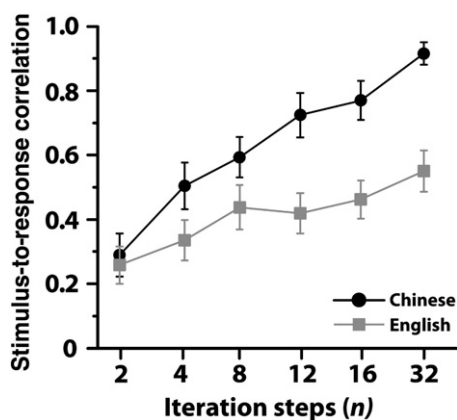
Fig. 2 – Group comparisons of FFR pitch-tracking accuracy in response to IRN homologues of Mandarin Tone 2 (T2) as a function of iteration steps. Symbols represent the group means of the stimulus-to-response correlation coefficients; error bars=±1 SE. Chinese show more faithful pitch tracking than English for iteration steps n≥12 only.

(group×section: $F_{5,90}=2.90$, $p=0.0178$; group×iteration: $F_{5,90}=4.15$, $p=0.0019$; iteration×section: $F_{25,450}=3.76$, $p<0.0001$). The three-way interaction failed to reach significance ($F_{25,450}=0.69$, $p=0.8684$). By section, post hoc Tukey–Kramer pairwise comparisons of group ($\alpha=0.05$) revealed that pitch strength was significantly greater in the native Chinese group as compared to the English in two sections of $T2_{IRN}$ that exhibit the most rapid changes in pitch (S4, S5; see Fig. 4) at almost all iteration steps ($n\geq4$).

As indexed by the exponential time constants (τ_C, τ_E), pitch strength emerges two (2.08) and three (2.98) times faster in Chinese than English listeners within sections 4 and 5, respectively. The changes in pitch within these two sections are characterized by higher rates of acceleration, i.e., pitch change per unit time expressed in semitones/ms (S4=0.0364, S5=0.0378) than the others (S1=−0.0095, S2=0.0, S3=0.0201, S6=0.0189). Thus, even in a case of degraded stimulus periodicity, we observe that FFRs of the Chinese are more sensitive to rapid changes of pitch *primarily* in those tonal sections exhibiting higher rates of acceleration.

3. Discussion

Using IRN homologues of Mandarin Tone 2 that vary in degree of pitch salience, the major findings of this cross-language study demonstrate that experience-dependent brainstem mechanisms for pitch representation are more sensitive in Chinese as compared to English listeners. Chinese, relative to English, show higher pitch-tracking accuracy of the whole tone except for the most severely degraded stimulus periodicities. Moreover, Chinese show greater pitch strength even in severely degraded stimuli for two 40-ms sections of $T2_{IRN}$ characterized by rapid changes in pitch. For these sections, exponential time constants reveal that pitch strength emerges much faster in the Chinese group as a function of increasing pitch salience than in the English group. These findings

altogether suggest that experience-dependent brainstem mechanisms for pitch are sensitive to specific dimensions of pitch contours that are highly relevant in degraded as well as normal listening conditions.

3.1. Experience-dependent plasticity of brainstem mechanisms underlying pitch extraction

Our findings lead us to infer that pitch extraction mechanisms in the brainstem of native speakers of a tonal language are less vulnerable to degradation of periodicity in IRN stimuli. The fact that language-dependent effects on pitch strength are restricted primarily to those sections of $T2_{IRN}$ that exhibit rapid changes in pitch (S4, S5; see Fig. 4) strongly suggests that neural plasticity in the brainstem is based on specific dimensions as opposed to the whole of a pitch contour. Moreover, time constants for these sections show that increasing pitch strength emerges at earlier iteration steps in Chinese listeners than in English. Because of its phonological system, native Chinese must be able to extract pitch relevant information, especially rapidly changing pitch glides within syllables, under adverse listening conditions. The relevance of rapid pitch changes can also be seen in speech production. First, it is observed that the maximum speed of pitch change that native Chinese are capable of is often approached in speech (Xu and Sun, 2002). Second, f_0 patterns in Mandarin have a greater amount of dynamic movement as a function of time and number of syllables than those in English (Eady, 1982). In terms of neural plasticity, our findings support the view that the extent to which noise interferes with hearing varies as a function of the perceptual relevance of particular dimensions of the speech signal. In the brainstem, extraction of these dimensions (e.g., rapidly changing pitch) is resilient to the negative effects of interfering noise. Such resilience would be ecologically useful to a native speaker especially in less than optimal listening conditions.

From the perspective of auditory neuroethology, we conclude that experience-driven adaptive neural mechanisms are involved subcortically, which sharpens response properties of neurons tuned for processing pitch contours that are sensitive to the prosodic needs of a particular language. This adjustment in processing pitch contours of Mandarin tones is comparable to neural mechanisms that are developed for processing behaviorally relevant sounds in other non-primate and non-human primate animals (Suga et al., 2003). Auditory processing is not limited to a simple representation of acoustic features of speech stimuli. Indeed, language-dependent operations begin well before the signal reaches the cerebral cortex.

3.2. Neural mechanisms underlying pitch encoding in the brainstem in response to degraded periodicities

The results of this study show that the brainstem pitch representation for the Chinese group, as compared to the English group, was more robust even when the temporal regularity of the stimulus was degraded so as to produce a weak pitch salience (as measured by FFR pitch strength) of the native tone. This finding suggests that long-term experience in a tonal language generates a pitch mechanism that not only
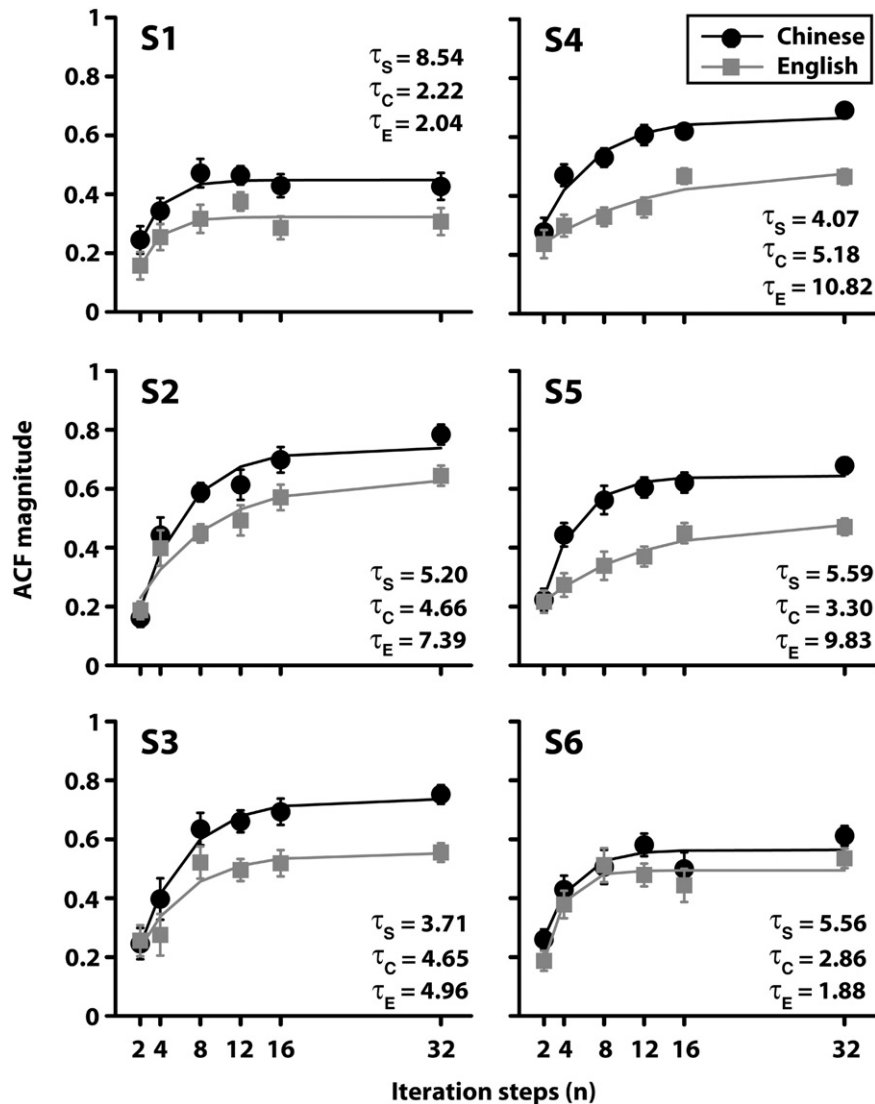
Fig. 3 – Group comparisons of pitch strength derived from FFR responses to T2$_{IRN}$ as a function of iteration steps (n). Individual panels show pitch strength computed within one of six contiguous 40 ms sections of the response (S1–S6; Fig. 3, top panel). FFR pitch strength of the Chinese group is greater than that of the English: S3 ($n \geq 12$); S4–S5 ($n \geq 4$). Solid lines show least-squares exponential fits to the data; error bars = ±1 SE. FFR pitch strength emerges more rapidly across iteration steps as reflected by the time constants for the Chinese (τ_C) and English (τ_E) groups (cf., τ_S, the stimulus time constant). The faster growth of neural pitch strength in the Chinese group, relative to the English, is especially evident in S4 and S5, those portions of T2 characterized by the most rapidly changing pitch (cf., **Fig. 4**, top panel).

is less susceptible to degradation of pitch relevant information in the stimulus but also is capable of extracting pitch information from patterns of neural activity where pitch related periodicities occur with low probability. We have previously proposed that this language-dependent enhancement of pitch representation reflects an enhanced tuning to interspike intervals that correspond especially to rapidly changing dynamic segments of the pitch trajectory (see Krishnan and Gandour, 2009, for review). This enhancement is likely mediated by temporal correlation analysis performed by coincidence detection neurons along the pitch axis (Langner, 1992, 2004), whose tuning characteristics are differentially sharpened as a function of experience. This sharpen-

ing is likely mediated by local excitatory and inhibitory interactions that are known to play an important role in signal selection at the level of the brainstem (Ananthanarayan and Gerken, 1983, 1987).

Such interaction may take the form of an active facilitation/disinhibition of the pitch intervals corresponding to the dynamic segments and inhibition of other pitch periods. It has been proposed that neuromodulatory inputs to the cortico-colliculo-thalamo-cortico-collicular loop influence the balance between excitation and inhibition (Xiong et al., 2009). In the brainstem, this same mechanism could possibly account for both the increased tolerance to signal degradation and the ability to extract pitch from relatively degraded pitch
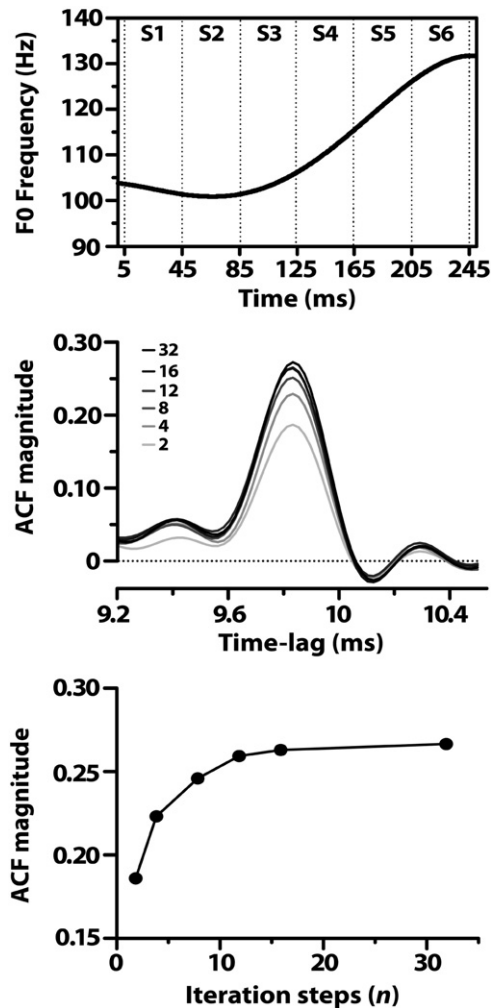
Fig. 4 – IRN stimuli used to evoke brainstem responses to a linguistic pitch pattern under varying degrees of temporal degradation. Voice fundamental frequency contour (f_0) of all six stimuli modeled after citation form Mandarin Tone 2 (T2) using a fourth-order polynomial equation (top panel). Six non-overlapping 40 ms sections (S1–S6) are demarcated by the vertical dotted lines. ACF magnitude around the pitch period of interest ($f_0 = 100$–135 Hz; period $= 7.4$–10 ms) increases with the number of iteration steps (n) used in the IRN generating circuit (middle panel). The height of the ACF peak increases exponentially with increasing n (bottom panel). ACF, autocorrelation function; IRN, iterated rippled noise.

relevant neural information at lower stimulus iterations. Specifically, the increased sensitivity of the autocorrelation analysis produces greater concentration of phase-locked intervals at the pitch period. Indeed, increases in pitch strength are positively correlated with increases in the number of intervals at the IRN delay (Patterson et al., 1996; Yost, 1996a). While the above explanation is more consistent with a local brainstem mechanism, we cannot rule out the possibility of top-down influence since brainstem neurons are subject to corticofugal modulation (Suga et al., 2000; Suga and Ma, 2003).

3.3. Do pitch representations in the brainstem reflect speech, language, lexical tone, music, or something else?

Albeit relevant to pitch representations in the brainstem, we claim that language, speech, music, or even lexical tone, are epiphenomenal. What is of direct relevance is native listeners' experience with pitch trajectories that commonly occur in a tone language in a wide variety of noisy environments. As a consequence, we observe enhanced sensitivity to linguistic pitch contours for even less than ideal auditory signals as compared to nonnative listeners. But these effects are not necessarily limited to an individual tone language and its tonal inventory, as enhancement of pitch extraction in the brainstem has been shown to transfer to other tone languages of similar typology (Krishnan et al., 2010). Nor can this experience-dependent enhancement of pitch representation of $T2_{IRN}$ be ascribed to a speech context (cf., Krishnan et al., 2009a,b). Although the origin for cross-language differences in FFR responses may emerge from exposure to natural speech, the experience-dependent effects are not circumscribed to speech perception (Xu et al., 2006a). Moreover, pitch extraction in the brainstem is not necessarily restricted by domain. FFRs to lexical tones may be enhanced by long-term experience with musical pitch patterns (Bidelman et al., in press; Wong et al., 2007) as well as by short-term training with second language learners (Song et al., 2008). Taken together, these findings imply that neural plasticity of brainstem responses is shaped by a listener's experience, and moreover, weighted by the relative saliency of *acoustic dimensions* underlying familiar pitch patterns.

Pitch-tracking accuracy of the whole contour of $T2_{IRN}$ is more accurate in the native than the nonnative group even for stimuli that have less than optimal pitch salience (e.g., $n = 12$; Fig. 2). The fact that experience-dependent neuroplasticity occurs even when the pitch contours are presented with degraded periodicity attests to the resilience of brainstem responses to linguistically relevant pitch. At $n = 32$, the upper limit of the continuum (Yost, 1996a), our finding of superior pitch tracking by the native group replicates earlier findings with Tone 2 in both speech (Krishnan et al., 2005) and nonspeech (Krishnan et al., 2009a,b; Swaminathan et al., 2008a) contexts. Yet pitch-tracking accuracy as a measure of robustness of pitch representations in the brainstem is insufficient in and of itself precisely because of its restriction to the *whole* of a pitch contour (Krishnan and Gandour, 2009).

Of special interest is the finding that cross-language differences in pitch strength occur in only in certain parts or *sections* of $T2_{IRN}$ (Fig. 3). In fact, pitch strength is greater in the native than in the nonnative group in only two sections (S4, S5) of $T2_{IRN}$ across all iterations except the lower limit ($n = 2$). This finding reinforces the view that neural mechanisms in the brainstem are not responding to lexical tone per se, but rather to specific time-varying acoustic features or dimensions of the auditory signal.

In the case of $T2_{IRN}$, the question is what are its time-varying dimensions that are crucially relevant to pitch processing in the brainstem? Our findings show that sections S4 and S5, as measured by the rate of acceleration, exhibit the most rapid changes in pitch within the whole of the f_0 contour. And not coincidentally, we observe that FFRs of the native

group are more sensitive than the nonnative in those same two sections that exhibit higher rates of acceleration. It is remarkable that these cross-language differences persist even in the context of severely degraded stimulus periodicity as early as $n=4$.

Thus, the rate of acceleration (and deceleration; cf. Krishnan et al., 2009b) of pitch trajectories appears to be a critical variable that influences pitch extraction in the rostral brainstem. We hypothesize that cross-language differences in the sustained phase-locked activity of the brainstem reflect an enhancement of selectivity to pitch relevant periodicities that correspond to rapidly changing dynamic portions of the pitch contour. The role of the brainstem is to facilitate cortical level processing of pitch relevant information by optimally capturing those features of the auditory signal that are of linguistic relevance.

3.4. Emergence of acoustic–phonetic features at subcortical stages of processing

Neural networks mediating lexical tone are likely to involve a series of computations that apply to representations at different levels of processing (Hickok and Poeppel, 2007; Poeppel et al., 2008). In the brainstem, FFR data on T2$_{IRN}$ gives us a window on the nature of features associated with pitch representations at an early sensory stage of processing. As reported herein, FFR responses are differentially sensitive to time-varying acceleration rates, i.e., changes in slope or overall shape of an f_0 contour, as a function of language experience.

Pitch itself, however, is a multidimensional perceptual attribute that relies on several acoustic features (e.g., height, direction, slope). The question is whether cross-language differences manifest at later perceptual stages of processing in the weighting of shape-related acoustic–phonetic features. At an *early, preattentive cortical* stage of pitch processing, a cross-language, mismatch negativity (MMN) study of Mandarin tones reveals two pitch dimensions derived from a multidimensional scaling analysis of MMN responses, interpretively labeled as "height" and "contour" (Chandrasekaran et al., 2007a,b). It is the contour dimension that best separates the Chinese and English individuals into their respective language group. In large-scale, cross-language multidimensional scaling studies of tone perception, three dimensions are reported to underlie a common tone space: height, direction, and slope (Gandour and Harshman, 1978; Gandour, 1983). The direction and slope (magnitude) dimensions are more effective than height in separating native tone language speakers, including Mandarin, from English. Thus, we see that cross-language effects on the saliency of acoustic–phonetic features of pitch percolate up the processing hierarchy beginning no later than 6–8 ms from the time the auditory signal enters the ear.

3.5. Conclusions

Background noise and/or competing sounds represent a significant challenge for neural mechanisms encoding behaviorally relevant acoustic features of the target sound. In the case of lexical tone, the extent to which a particular feature of pitch is resistant to degraded listening conditions

may serve as an index of its perceptual saliency. How pitch information is extracted in both subcortical and cortical areas depends on the interactions between specific features of the input signal, their corresponding output representations, and the nature of pitch experience of the listener (cf., Zatorre, 2008, p. 533). Therefore, it is critical that experience-enhanced sensitivity of brainstem neural mechanisms allow extraction and preservation of information about acoustic features that are important for speech perception. The findings herein represent such an adaptation of the brainstem pitch mechanism(s) to encode and transmit robust pitch relevant information to higher centers even under conditions of stimulus degradation.

4. Experimental procedures

4.1. Participants

Ten adult native speakers of Mandarin Chinese (5 males, 5 females), hereafter referred to as Chinese (C), and 10 adult native speakers of American English (4 males, 6 females), referred to as English (E), participated in the FFR experiment. The two groups were closely matched in age (Chinese: mean=25.5, SD=2.5; English: mean=23.8, SD=3.4), education (Chinese: mean=18.5, SD=1.4; English: mean=17.3, SD=2.2), and were strongly right handed (>84.7 %) as measured by the Edinburgh Handedness Inventory (Oldfield, 1971). All participants exhibited normal hearing sensitivity (better than 20 dB HL in both ears) at octave frequencies from 500 to 4000 Hz. In addition, participants reported no previous history of neurological or psychiatric illnesses. Each participant completed a language history questionnaire (Li et al., 2006). Native speakers of Mandarin were born and raised in mainland China and none received formal instruction in English before the age of 9 (mean=12.2, SD=1.6). The English-speaking group had no prior experience learning any tonal language whatsoever. Participants also completed a music history questionnaire (Wong and Perrachione, 2007). None had more than 3 years of formal music training on any combination of instruments nor had studied music within the past 5 years. All participants were students enrolled at Purdue University at the time of their participation. All were paid for their time and gave informed consent in compliance with a protocol approved by the Institutional Review Board of Purdue University.

4.2. Stimuli

A stimulus continuum varying in pitch salience (from low to high) was created by systematically increasing the number of iterations of IRN homologues of a lexical tone contour. Instead of a single static delay, time-varying delays were used to create IRN with a dynamic contour whose pitch varies as a function of time (Krishnan et al., 2009b; Swaminathan et al., 2008b). The fundamental frequency (f_0) contour of each stimulus was modeled after the natural citation form of Mandarin Tone 2 (T2), as produced by a male speaker (Xu, 1997), using a fourth-order polynomial (Fig. 4, top panel). All stimuli were matched in RMS amplitude and

fixed in duration to 250 ms including a 10-ms rise/fall time (cos^2 ramps). The gain in the IRN circuit was set to 1 for all the stimuli.

T2$_{IRN}$ was chosen primarily because it elicits differential sensitivity to time-varying acceleration rates when presented using an *optimal* number of iteration steps (Bidelman et al., in press; Krishnan et al., 2009a,b). By using this same stimulus, we were able to determine whether this language-dependent differential sensitivity to time-varying acceleration rates is maintained in response to *less than optimal* degraded periodicities extending downward toward the noise end of the continuum.

Six versions of T2 were created using this same pitch contour at iteration steps of $n = 2, 4, 8, 12, 16,$ and 32 (Supplementary Material: audio files, .mp3; IRNstimuli.doc). The lower limit of $n = 2$ is equivalent to noise. The upper limit of $n = 32$ was chosen because pitch strength does not increase by any noticeable amount beyond this number of iteration steps (Yost, 1996a). Intermediate iteration steps were chosen based on results from extensive pilot work in which we determined that this continuum yielded measurable differences in brainstem responses.

In these IRN stimuli, the magnitude of the autocorrelation function (ACF) around the pitch period of interest ($f_0 = 100$–135 Hz; period = 7.4–10 ms) increases systematically with increasing n (Fig. 4, middle panel). Both behavioral (Yost, 1996a) and neurophysiological (Cariani and Delgutte, 1996a,b) studies have shown a high correspondence between the ACF peak height (Fig. 4, bottom panel) and the growth of perceived pitch salience.

4.3. Data acquisition

FFR recording protocol and data analysis are similar to those reported in previous publications from our laboratory (Krishnan and Gandour, 2009; Krishnan et al., 2009a,b).

Participants reclined comfortably in an acoustically and electrically shielded booth. They were instructed to relax and refrain from extraneous body movements to minimize myogenic artifacts. FFRs were recorded from each participant in response to monaural stimulation of the right ear at 80 dB SPL using a repetition rate of 2.76/s. The presentation order of the stimuli was randomized both within and across participants. Control of the experimental protocol was accomplished by a signal generation and data acquisition system (Tucker-Davis Technologies, System III). The stimulus files were routed through a digital to analog module and presented through a magnetically shielded insert earphone (Etymotic, ER-3A).

FFRs were recorded differentially between a non-inverting (positive) electrode placed on the midline of the forehead at the hairline (Fz) and inverting (reference) electrodes placed on (i) the right mastoid (A2), (ii) the left mastoid (A1), and (iii) the 7th cervical verterbra (C7). Another electrode placed on the mid-forehead (Fpz) served as the common ground. FFRs were recorded simultaneously from the three different electrode configurations and subsequently averaged for each stimulus condition to yield a response with a higher signal-to-noise ratio (Krishnan et al., 2009b). All interelectrode impedances were maintained at or below 1 kΩ. The raw EEG inputs were amplified by 200,000 and band-pass filtered from 80 to 3000 Hz (6 dB/octave roll-off, RC response characteristics). Each FFR waveform represents the average of 3000 stimulus presentations over a 280-ms analysis window using a sampling rate of 24,414 kHz.

4.4. Data analysis

4.4.1. Temporal analysis

Short-term ACFs and running autocorrelograms were computed for the grand averaged FFRs to index variation in FFR periodicities over the duration of the response. The autocorrelogram (ACG) represents the short-term autocorrelation function of windowed frames of a compound signal, i.e., ACG $(\tau, t) = X(t) \times X(t - \tau)$ for each time t and time-lag τ. It is a three-dimensional plot quantifying the variations in periodicity and pitch strength as a function of time. The horizontal axis represents the time at which single ACF "slices" are computed while the vertical axis represents their corresponding time-lags, i.e., pitch periods. The intensity of each point in the image represents the instantaneous ACF magnitude computed at a given time within the response. Mathematically, the running autocorrelogram is the time–domain analog to the frequency–domain spectrogram. In terms of neurophysiology, it represents the running distribution of all-order interspike intervals present in the population neural activity (Cariani and Delgutte, 1996b).

4.4.2. Pitch-tracking accuracy of whole stimuli

The ability of the FFR to follow pitch changes in the stimuli was evaluated by extracting the f_0 contour from the FFRs using a periodicity detection short-term autocorrelation algorithm (Boersma, 1993). Essentially, the algorithm works by sliding a 40-ms window in 10-ms increments over the time course of the FFR. The autocorrelation function was computed for each 40-ms frame and the time-lag corresponding to the maximum autocorrelation value within each frame was recorded. The reciprocal of this time-lag, or pitch period, represents an estimate of f_0. The time-lags associated with autocorrelation peaks from each frame were concatenated together to give a running f_0 contour. This analysis was performed on both the FFRs and their corresponding stimuli. Pitch-tracking accuracy is computed as the cross-correlation coefficient between the f_0 contour extracted from the FFRs and the f_0 contour extracted from the stimulus.

4.4.3. Pitch strength of stimuli sections

To compute the pitch strength of the FFRs to *time-varying* stimuli, responses were divided into six non-overlapping 40-ms sections: S1, 5–45; S2, 45–85; S3, 85–125; S4, 125–165; S5, 165–205; S6, 205–245 ms (Fig. 4). The normalized autocorrelation function (expressed as a value between 0 and 1) was computed for each of these sections, where 0 represents an absence of periodicity and 1 represents maximal periodicity. Within each 40-ms section, a response peak was selected which corresponded to the same location (time-lag) of the autocorrelation peak in the input stimulus (Swaminathan et al., 2008a). The magnitude of this response peak represents an estimate of the pitch strength per section. All data analyses were performed using custom

routines coded in MATLAB® 7.5 (The MathWorks, Inc., Natick, MA).

4.5. Statistical analysis

Pitch-tracking accuracy was measured as the cross-correlation coefficient between the f_0 contours extracted from T2$_{IRN}$ (stimulus) and the f_0 contours extracted from the FFRs (response) at each iteration step. A mixed model ANOVA (SAS®) with subjects as a random factor nested within group (C, E), the between-subject factor, and iteration step (2, 4, 8, 12, 16, 32), the within-subject factor, was conducted on the cross-correlation coefficients to evaluate the effects of language experience on the ability of the FFR to track f_0 contours under varying degrees of temporal degradation.

Pitch strength (magnitude of the normalized autocorrelation peak) was calculated for each of the six sections within T2$_{IRN}$ at each iteration step for every subject. These pitch strength values were analyzed using a three-way ANOVA with subjects as a random factor nested within group (C, E), and with iteration steps (2, 4, 8, 12, 16, 32) and section (S1, S2, S3, S4, S5, S6) as fully crossed within-subject factors. By examining iteration steps, we were able to assess whether cross-language effects on pitch strength vary depending on the degree of temporal degradation. By examining 40-ms sections, we were able to evaluate whether cross-language effects on pitch strength are restricted to specific portions of the f_0 contour, and moreover, whether those portions are differentially affected by the degree of temporal degradation, or pitch salience.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.brainres.2009.11.061.

R E F E R E N C E S

Alkhoun, I., Gallégo, S., Moulin, A., Ménard, M., Veuillet, E., Berger-Vachon, C., Collet, L., Thai-Van, H., 2008. The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. Clin. Neurophysiol. 119, 922–933.

Ananthanarayan, A.K., Gerken, G.M., 1983. Post-stimulation effects on the auditory brain stem response partial-masking and enhancement. Electroencephalogr. Clin. Neurophysiol. 55, 223–226.

Ananthanarayan, A.K., Gerken, G.M., 1987. Response enhancement and reduction of the auditory brain-stem response in a forward-masking paradigm. Electroencephalogr. Clin. Neurophysiol. 66, 427–439.

Bidelman, G.M., Gandour, J.T., Krishnan, A., in press. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. J. Cogn. Neurosci. (posted online 19 Nov 2009). doi:10.1162/jocn.2009.21362.

Bilsen, F.A., ten Kate, J.H., Buunen, T.J.F., Raatgever, J., 1975. Responses of single units in the cochlear nucleus of the cat to cosine noise. J. Acoust. Soc. Am. 58, 858–866.

Boersma, P., 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proc. Inst. Phon. Sci. 17, 97–110.

Cariani, P.A., Delgutte, B., 1996a. Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. J. Neurophysiol. 76, 1717–1734.

Cariani, P.A., Delgutte, B., 1996b. Neural correlates of the pitch of complex tones: I. Pitch and pitch salience. J. Neurophysiol. 76, 1698–1716.

Chandrasekaran, B., Gandour, J.T., Krishnan, A., 2007a. Neuroplasticity in the processing of pitch dimensions: a multidimensional scaling analysis of the mismatch negativity. Restor. Neurol. Neurosci. 25, 195–210.

Chandrasekaran, B., Krishnan, A., Gandour, J.T., 2007b. Mismatch negativity to pitch contours is influenced by language experience. Brain Res. 1128, 148–156.

Eady, S.J., 1982. Differences in the F_0 patterns of speech: tone language versus stress language. Lang. Speech 25, 29–42.

Fay, R.R., Yost, W.A., Coombs, S., 1983. Psychophysics and neurophysiology of repetition noise processing in a vertebrate auditory system. Hear. Res. 12, 31–55.

Galbraith, G.C., Threadgill, M., Hemsley, J., Salour, K., Songdej, N., Ton, J., Cheung, L., 2000. Putative measure of peripheral and brainstem frequency-following in humans. Neurosci. Lett. 292, 123–127.

Gandour, J.T., 1983. Tone perception in Far Eastern languages. J. Phon. 11, 149–175.

Gandour, J.T., Harshman, R.A., 1978. Crosslanguage differences in tone perception: a multidimensional scaling investigation. Lang. Speech 21, 1–33.

Glaser, E.M., Suter, C.M., Dasheiff, R., Goldberg, A., 1976. The human frequency-following response: its behavior during continuous tone and tone burst stimulation. Electroencephalogr. Clin. Neurophysiol. 40, 25–32.

Griffiths, T.D., Buchel, C., Frackowiak, R.S., Patterson, R.D., 1998. Analysis of temporal structure in sound by the human brain. Nat. Neurosci. 1, 422–427.

Griffiths, T.D., Uppenkamp, S., Johnsrude, I., Josephs, O., Patterson, R.D., 2001. Encoding of the temporal regularity of sound in the human brainstem. Nat. Neurosci. 4, 633–637.

Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402.

Johnson, K.L., Nicol, T.G., Kraus, N., 2005. Brain stem response to speech: a biological marker of auditory processing. Ear Hear. 26, 424–434.

Kraus, N., Banai, K., 2007. Auditory-processing malleability: focus on language and music. Curr. Dir. Psychol. Sci. 16, 105–110.

Kraus, N., Nicol, T., 2005. Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. Trends Neurosci. 28, 176–181.

Krishnan, A., 2006. Human frequency following response. In: Burkard, R.F., Don, M., Eggermont, J.J. (Eds.), Auditory Evoked Potentials: Basic Principles and Clinical Application. Lippincott Williams & Wilkins, Baltimore, pp. 313–335. Vol.

Krishnan, A., Gandour, J.T., 2009. The role of the auditory brainstem in processing linguistically-relevant pitch patterns. Brain Lang. 110, 135–148.

Krishnan, A., Xu, Y., Gandour, J., Cariani, P., 2005. Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res. Cogn. Brain Res. 25, 161–168.

Krishnan, A., Gandour, J.T., Bidelman, G.M., Swaminathan, J., 2009a. Experience-dependent neural representation of dynamic pitch in the brainstem. Neuroreport 20, 408–413.

Krishnan, A., Swaminathan, J., Gandour, J.T., 2009b. Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. J. Cogn. Neurosci. 21, 1092–1105.

Krishnan, A., Gandour, J.T., Bidelman, G.M., 2010. The effects of tone language experience on pitch processing in the brainstem. J. Neurolinguistics. 23, 81–95.

Krumbholz, K., Patterson, R.D., Seither-Preisler, A., Lammertmann, C., Lutkenhoner, B., 2003. Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. Cereb. Cortex 13, 765–772.

Langner, G., 1992. Periodicity coding in the auditory system. Hear. Res. 60, 115–142.

Langner, G., 2004. Topographic representation of periodicity information: the 2nd neural axis of the auditory system. In: Syka, J., Merzenich, M. (Eds.), Plasticity of the Central Auditory System and Processing of Complex Acoustic Signals. Plenum Press, New York, pp. 21–26. Vol.

Li, P., Sepanski, S., Zhao, X., 2006. Language history questionnaire: a Web-based interface for bilingual research. Behav. Res. Methods 38, 202–210.

Marsh, J.T., Brown, W.S., Smith, J.C., 1974. Differential brainstem pathways for the conduction of auditory frequency-following responses. Electroencephalogr. Clin. Neurophysiol. 36, 415–424.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

Patel, A.D., Iversen, J.R., 2007. The linguistic benefits of musical abilities. Trends Cogn. Sci. 11, 369–372.

Patterson, R.D., Handel, S., Yost, W.A., Datta, A.J., 1996. The relative strength of the tone and noise components in iterated ripple noise. J. Acoust. Soc. Am. 100, 3286–3294.

Poeppel, D., Idsardi, W.J., van Wassenhove, V., 2008. Speech perception at the interface of neurobiology and linguistics. Philos. Trans. R. Soc. Lond., B. Biol. Sci. 363, 1071–1086.

Sayles, M., Winter, I.M., 2007. The temporal representation of the delay of dynamic iterated rippled noise with positive and negative gain by single units in the ventral cochlear nucleus. Brain Res 1171, 52–66.

Shofner, W.P., 1991. Temporal representation of rippled noise in the anteroventral cochlear nucleus of the chinchilla. J. Acoust. Soc. Am. 90, 2450–2466.

Shofner, W.P., 1999. Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: analysis of neural autocorrelograms. J. Neurophysiol. 81, 2662–2674.

Smith, J.C., Marsh, J.T., Brown, W.S., 1975. Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. Electroencephalogr. Clin. Neurophysiol. 39, 465–472.

Soeta, Y., Nakagawa, S., Tonoike, M., 2005. Auditory evoked magnetic fields in relation to iterated rippled noise. Hear. Res. 205, 256–261.

Song, J.H., Skoe, E., Wong, P.C.M., Kraus, N., 2008. Plasticity in the adult human auditory brainstem following short-term linguistic training. J. Cogn. Neurosci. 20, 1892–1902.

Suga, N., Ma, X., 2003. Multiparametric corticofugal modulation and plasticity in the auditory system. Nat. Rev. Neurosci. 4, 783–794.

Suga, N., Gao, E., Zhang, Y., Ma, X., Olsen, J.F., 2000. The corticofugal system for hearing: recent progress. Proc. Natl. Acad. Sci. U. S. A. 97, 11807–11814.

Suga, N., Ma, X., Gao, E., Sakai, M., Chowdhury, S.A., 2003. Descending system and plasticity for auditory signal processing: neuroethological data for speech scientists. Sp. Comm. 41, 189–200.

Swaminathan, J., Krishnan, A., Gandour, J.T., 2008a. Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. Neuroreport 19, 1163–1167.

Swaminathan, J., Krishnan, A., Gandour, J.T., Xu, Y., 2008b. Applications of static and dynamic iterated rippled noise to evaluate pitch encoding in the human auditory brainstem. IEEE Trans. Biomed. Eng. 55, 281–287.

Tzounopoulos, T., Kraus, N., 2009. Learning to encode timing: mechanisms of plasticity in the auditory brainstem. Neuron 62, 463–469.

Winter, I.M., Wiegrebe, L., Patterson, R.D., 2001. The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea-pig. J. Physiol. 537, 553–566.

Winter, I.M., Palmer, A.R., Wiegrebe, L., Patterson, R.D., 2003. Temporal coding of the pitch of complex sounds by presumed multipolar cell in the ventral cochlear nucleus. Sp. Comm. 41, 135–149.

Wong, P.C., Perrachione, T.K., 2007. Learning pitch patterns in lexical identification by native English-speaking adults. Appl. Psycholinguist. 28, 565–585.

Wong, P.C., Skoe, E., Russo, N.M., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat. Neurosci. 10, 420–422.

Worden, F.G., Marsh, J.T., 1968. Frequency-following (microphonic-like) neural responses evoked by sound. Electroencephalogr. Clin. Neurophysiol. 25, 42–52.

Xiong, Y., Zhang, Y., Yan, J., 2009. The neurobiology of sound-specific auditory plasticity: A core neural circuit. Neuroscience and Biobehavioral Reviews 33, 1178–1184.

Xu, Y., 1997. Contextual tonal variations in Mandarin. Jphon. 25, 61–83.

Xu, Y., Sun, X., 2002. Maximum speed of pitch change and how it may relate to speech. J. Acoust. Soc. Am. 111, 1399–1413.

Xu, Y., Gandour, J.T., Francis, A.L., 2006a. Effects of language experience and stimulus complexity on the categorical perception of pitch direction. J. Acoust. Soc. Am. 120, 1063–1074.

Xu, Y., Krishnan, A., Gandour, J.T., 2006b. Specificity of experience-dependent pitch representation in the brainstem. Neuroreport 17, 1601–1605.

Yip, M., 2003. Tone, Vol. Cambridge University Press, New York.

Yost, W.A., 1996a. Pitch strength of iterated rippled noise. J. Acoust. Soc. Am. 100, 3329–3335.

Yost, W.A., 1996b. Pitch of iterated rippled noise. J. Acoust. Soc. Am. 100, 511–518.

Yost, W.A., 1997. Pitch strength of iterated rippled noise when the pitch is ambiguous. J. Acoust. Soc. Am. 101, 1644–1648.

Zatorre, R.J., 2008. Musically speaking. Neuron 26, 532–533.

Zatorre, R.J., Gandour, J.T., 2008. Neural specializations for speech and pitch: moving beyond the dichotomies. Philos. Trans. R. Soc. Lond., B. Biol. Sci. 363, 1087–1104.